1 DISTRIBUCIONES CONTINUAS.

2 CÁLCULO DE PROBABILIDADES.

- **Ejemplo 1:** Una persona informal hace esperar a su pareja aleatoriamente entre 0 y 90 minutos. Harto de esta situación, la persona que sufre la espera se plantea un ultimátum; sí al día siguiente su pareja tarda menos de 15 minutos mantiene la relación, sí la espera está entre 15 y 55 minutos, decide en la siguiente cita con los mismos criterios, mientras que si tarda más de 55 minutos la relación termina en ese momento.

a) Calcule la probabilidad de que la relación continúe hasta la siguiente cita.

```
x <- 55
a=0
b = 90
#usando la función propia de R
punif(x, min=a, max=b, lower.tail=TRUE)

## [1] 0.6111111
```

b) Calcule la probabilidad de que la relación termine en la segunda cita. Suponiendo que el tiempo de espera en una cita es independiente respecto de otras citas, se calcula la probabilidad $P(15 < X < 55) = P(X < 55) - P(X \leq 15) = 0.6111 - 0.1666 = 0.4445$

```
F55=punif(55, min=a, max=b, lower.tail=TRUE)
F15=punif(15, min=a, max=b, lower.tail=TRUE)
F55-F15

## [1] 0.4444444
```

que es la probabilidad de que aplace la decisión para la segunda cita y, en la segunda cita, la probabilidad de que lo deje definitivamente es $P(X > 55) = 0.3888$

```
F55=punif(55, min=a, max=b, lower.tail=TRUE)
F55

## [1] 0.6111111
```

luego multiplicando ambas probabilidades se obtiene el valor pedido 0.1728.

```
(1-F55)*( F55-F15)

## [1] 0.1728395
```

- **Ejemplo 2:** Una empresa está buscando personal para su departamento de mercadeo. El perfil solicitado es el de sujetos extrovertidos y creativos. Se han presentado 50 candidatos y la empresa ha establecido como criterio de selección que los candidatos superen el percentil 80 en creatividad y extroversión. Sabiendo que la variable extroversión (X) se distribuye según una Normal de media 5 y desviación típica 1, que la variable creatividad (Y) sigue una t-Student de 10 grados de libertad y que las puntuaciones de creatividad y extroversión son independientes:

a) ¿Cuántos candidatos serán seleccionados?
Al ser X e Y independientes, la probabilidad $PX \geq P80 \cap Y \geq P80 = P(X \geq P80)P(Y \geq P80) = (0.20)(0.20) = 0.04$ Como se han presentado 50 aspirantes, serán seleccionadas $(50)(0.04) = 2$ personas.

b) ¿Qué puntuaciones debe superar un aspirante en creatividad y extroversión para ser admitido? Según el criterio de selección se debe superar el percentil 80, en ambas variables, para ser admitido. Se calculará pues el percentil 80 de la variable X e Y, utilizando los cuantiles-normales para la variable X:

```
#y los cuantiles-normales para la variable X:
p <- c(0.80)
media=5
d.t=1
qnorm(p, mean=media, sd=d.t, lower.tail=TRUE)
```

```
## [1] 5.841621

#y los cuantiles-t para la variable Y:
p <- c(0.80)
g.l <- 10
qt(p, df=g.l, lower.tail=TRUE)

## [1] 0.8790578
```

c) Si se extraen al azar 16 candidatos, ¿cuál es la probabilidad de que su media aritmética en extroversión sea mayor que 4.5?
Se sabe que al extraer una muestra de una población normal de tamaño n, la media muestral, sigue otra distribución normal de media igual que la poblacional y desviación típica $\sigma/\sqrt{n}$

Como se desea calcular $P(\bar{x} \geq 4.5)$

```
n <- 16
x <- 4.5
mu=5
sigma=1
d.t=sigma/sqrt(n)
pnorm(x, mean=mu, sd=d.t, lower.tail=FALSE)

## [1] 0.9772499
```

- **Ejemplo 3:** La duración media de un modelo de marcapasos es de 7 años.

a) ¿Cuál es la probabilidad de que dure al menos 5 años? ¿y menos de 3 años? Suponiendo que la variable X="tiempo de funcionamiento del marcapasos" sigue una distribución exponencial con parámetro $\lambda = 1/\theta = 1/7$ con $\theta = E[X]$ tiempo promedio.

La probabilidad $P(X \geq 5)$ se obtiene así:

```
x <- 5
teta=7
pexp(x, rate=1/teta, lower.tail=FALSE)

## [1] 0.4895417
```

y de igual forma $P(X < 3)$:

```
x <- 3
teta=7
pexp(x, rate=1/teta, lower.tail=TRUE)

## [1] 0.3485609
```

b) Si han transcurrido ya 4 años desde su implantación, ¿cuál es la probabilidad de que dure otros 4?
Nos piden $P(X \geq 8/X \geq 4)$ Teniendo en cuenta que la función de distribución es la única distribución continua no tiene memoria resulta que $P(X \geq 8/X \geq 4) = P(X \geq 4) = 0.5647182$

```
pexp(4, rate=1/teta, lower.tail=FALSE)

## [1] 0.5647181
```

c) ¿Cuánto tiempo debería funcionar un marcapasos para estar entre el 10% de los que más duran?

Hay que calcular el percentil 90:

```
p <- 0.9
teta <- 7
qexp(p, rate=1/teta, lower.tail=TRUE)

## [1] 16.1181

#resultando 16.12 años.
```

d) Calcular el valor que deben tener a y b para que $P(X < a) = 0.5$ y $P(X > b) = 0.32$
De forma análoga al apartado anterior, en el primer caso habría que calcular la mediana (percentil 50), $a = 4.852$,

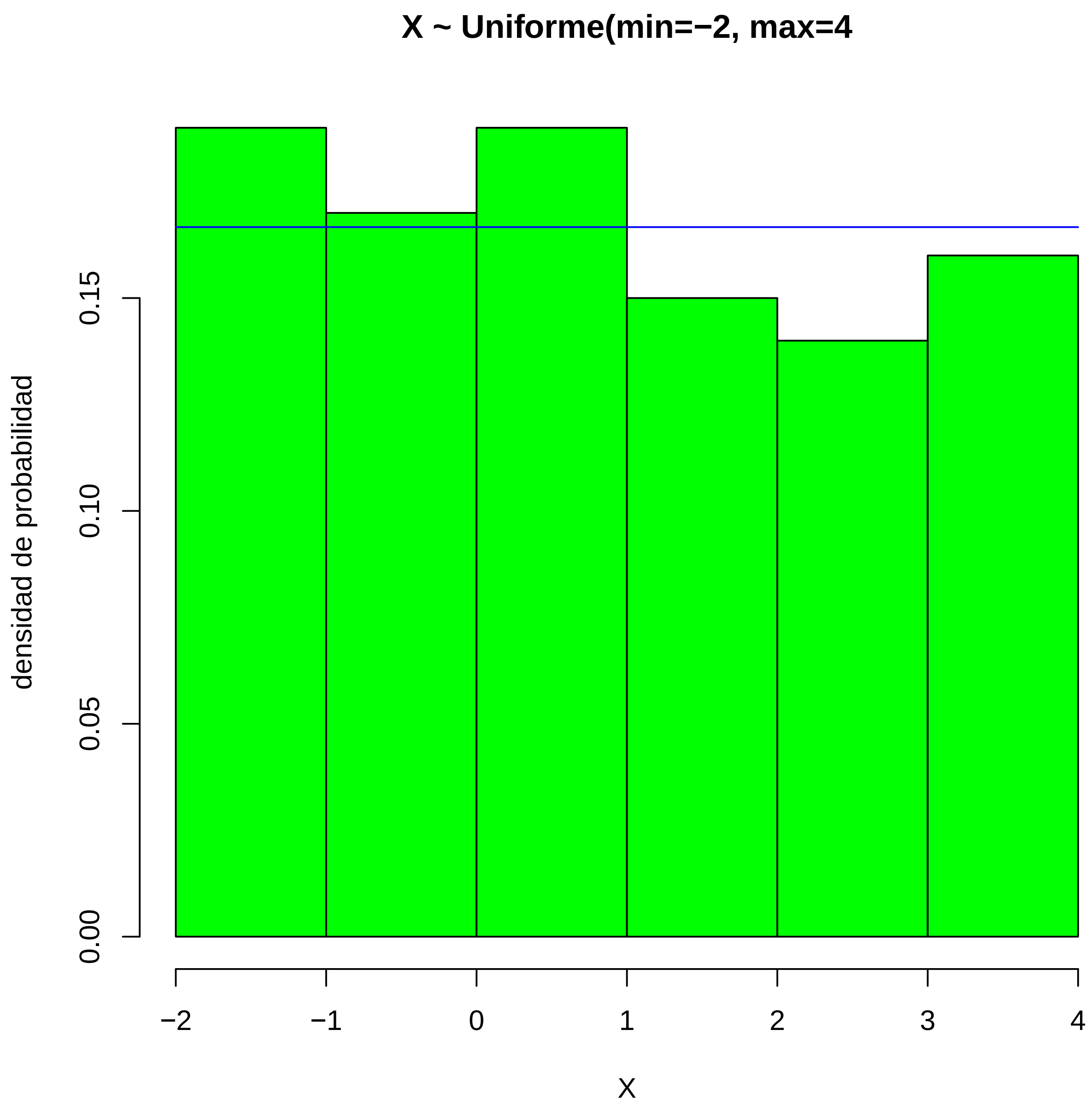
```
qexp(0.5, rate=1/teta, lower.tail=TRUE)

## [1] 4.85203

#y en el segundo caso, el percentil 68, b = 7.97
qexp(0.68, rate=1/teta, lower.tail=TRUE)

## [1] 7.97604

#o de esta otra manera
qexp(0.32, rate=1/teta, lower.tail=FALSE)

## [1] 7.97604
```

3 GENERACIÓN DE MUESTRAS ALEATORIAS DE LAS DISTRIBUCIONES

- **Ejemplo 1:**

Generar 100 números aleatorios de una distribución Uniforme en [-2, 4]

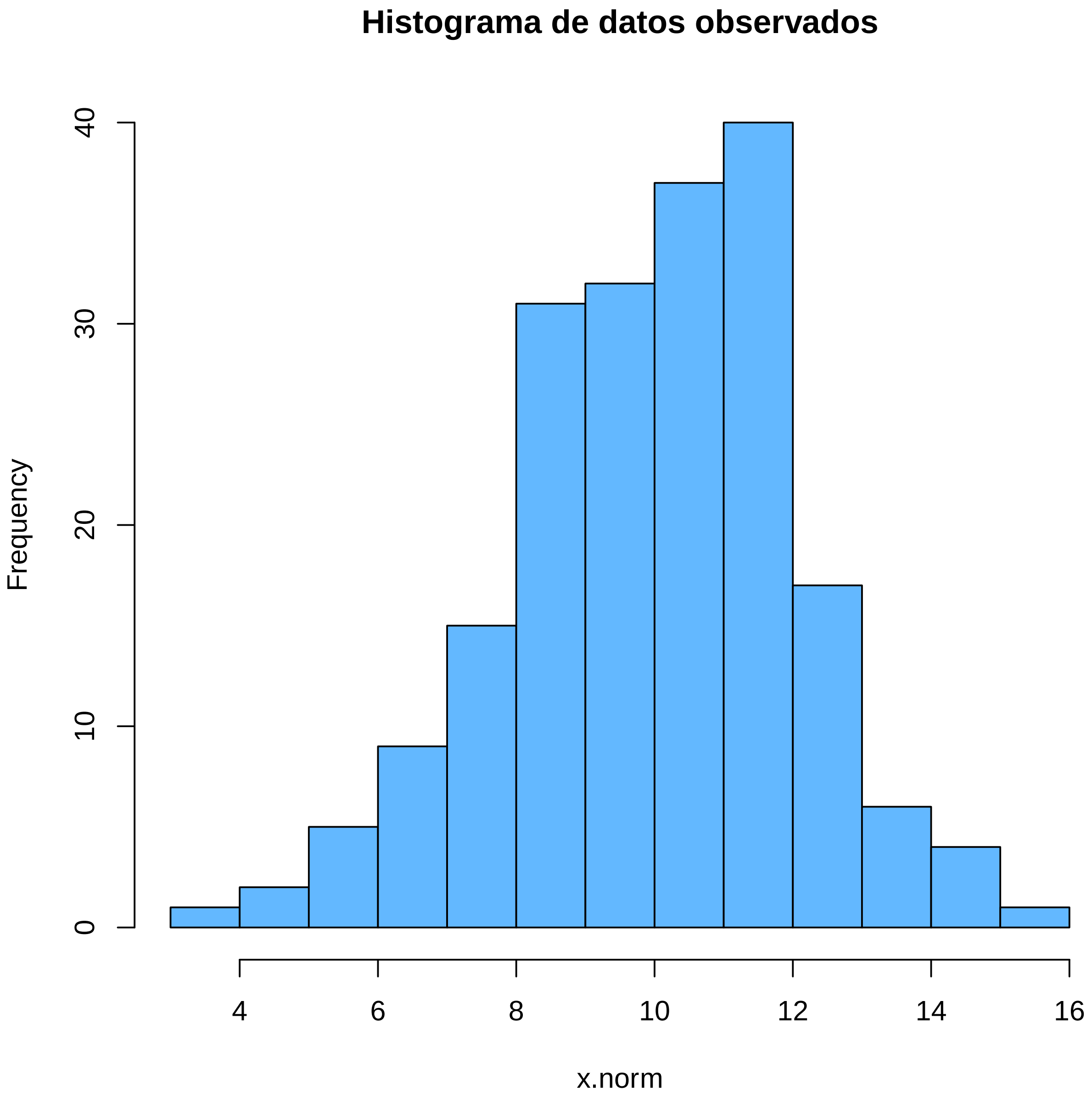
```
# Definir los parámetros apropiados
min <- -2
max <- 4

# generar 100 números aleatorios de la distribución
x = runif(100, min, max)
x

##   [1] -0.9558970  3.1267132  2.9301148  3.0728663 -1.9996469 -1.1575723
##   [7] -0.8175966  3.6627671  1.1268875  1.5073681  0.3637713  1.4771294
##  [13]  1.8240445  3.3556542  3.8575335 -0.2240745 -0.6841308 -0.7592002
##  [19]  0.1786235  1.4772483  1.3853160 -1.1310822  1.0070083  0.6958662
##  [25]  0.3940913  0.3496384 -1.0748612  0.9419767  2.4367016  0.9230461
##  [31]  3.4263482 -1.8085061 -0.8656259 -1.0756212 -1.3181165  0.7996228
##  [37] -0.6461161  3.9447520 -1.1025909 -1.3139665 -1.5999980 -1.7944386
##  [43]  2.4709718  3.2035791  0.6778555 -1.5965019  2.2808747  2.2244799
##  [49]  1.6444735 -0.7042938 -1.4373954  2.4726643  2.4619100 -0.8913438
##  [55] -0.3033126  1.6922182  3.5525729  0.8675269  1.6188031  3.2418490
##  [61]  1.3755856  1.7072274  2.5742426 -0.1015215  0.3161341  1.0259765
##  [67]  1.5331821  3.5986910  2.4633124 -0.4980027 -0.9888037  2.9070235
##  [73]  2.3505369 -1.5609153  2.6937089  0.6912957  3.4582293  0.5389910
##  [79]  2.4411304 -0.2387290 -1.0195454  0.3386082  3.3638813  2.0243261
##  [85]  0.1800639 -1.6897113  0.5429544  0.3912573 -0.8895319  3.0423491
##  [91]  3.6695666  0.6288016  3.5103912 -0.3011281  0.1826353 -1.1905739
##  [97]  1.5482364 -0.1088070 -1.8932998 -1.0458823

# Histograma para la nuestra aleatoria de tamaño 100
hist(x, main="X ~ Uniforme(min=-2, max=4", xlab="X", ylab="densidad de probabilidad",
     probability=TRUE, col="green")

# Graficar la función de densidad, use la función curve() para variable continua
curve(dunif(x, min, max), col="blue", add=TRUE)
```



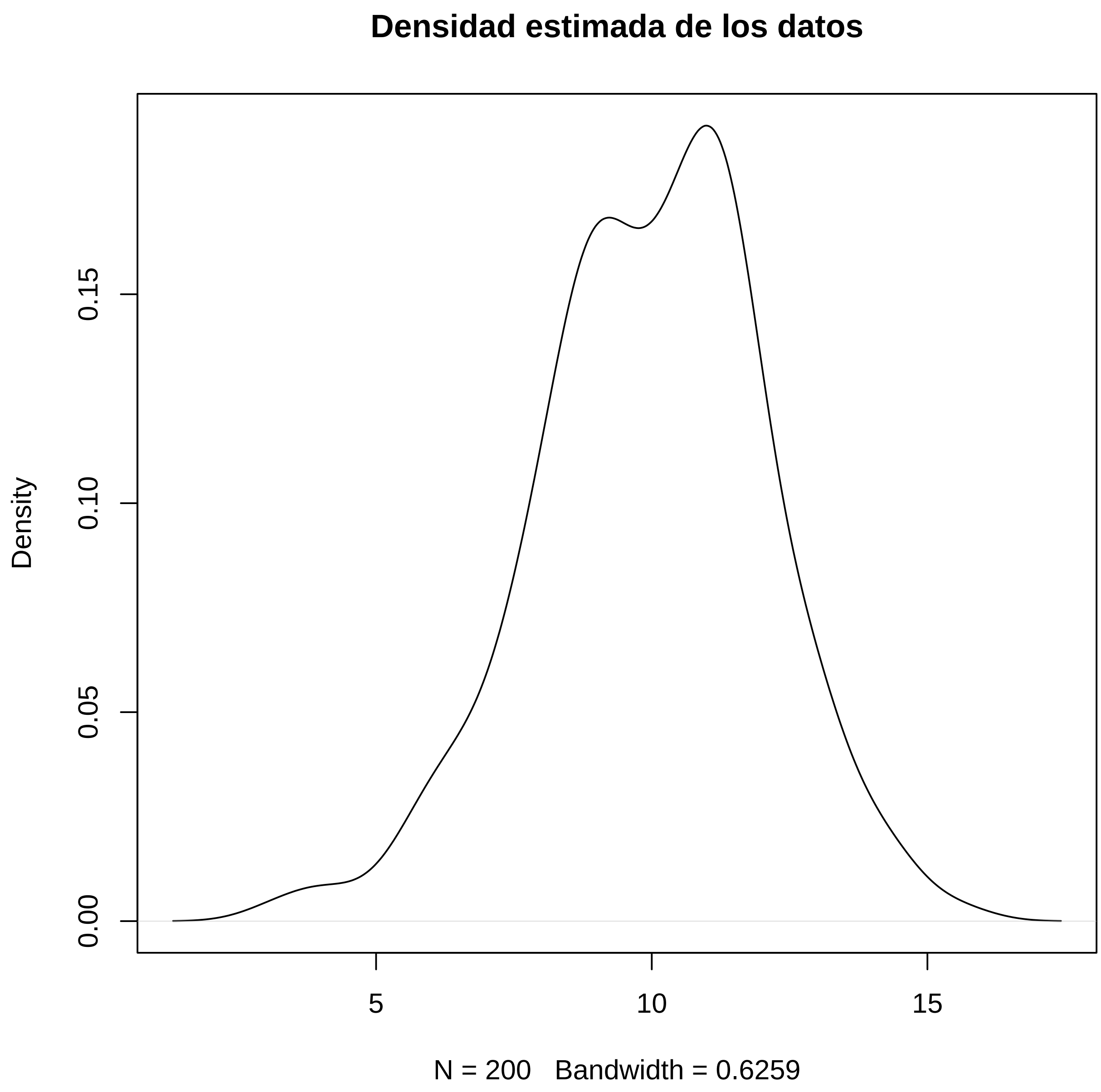
- **Ejemplo 2:**

Supongamos que tenemos una muestra de tamaño n=200 perteneciente a una población normal $N(10, 2)$ con $\mu = 10$ y $\sigma = 2$:

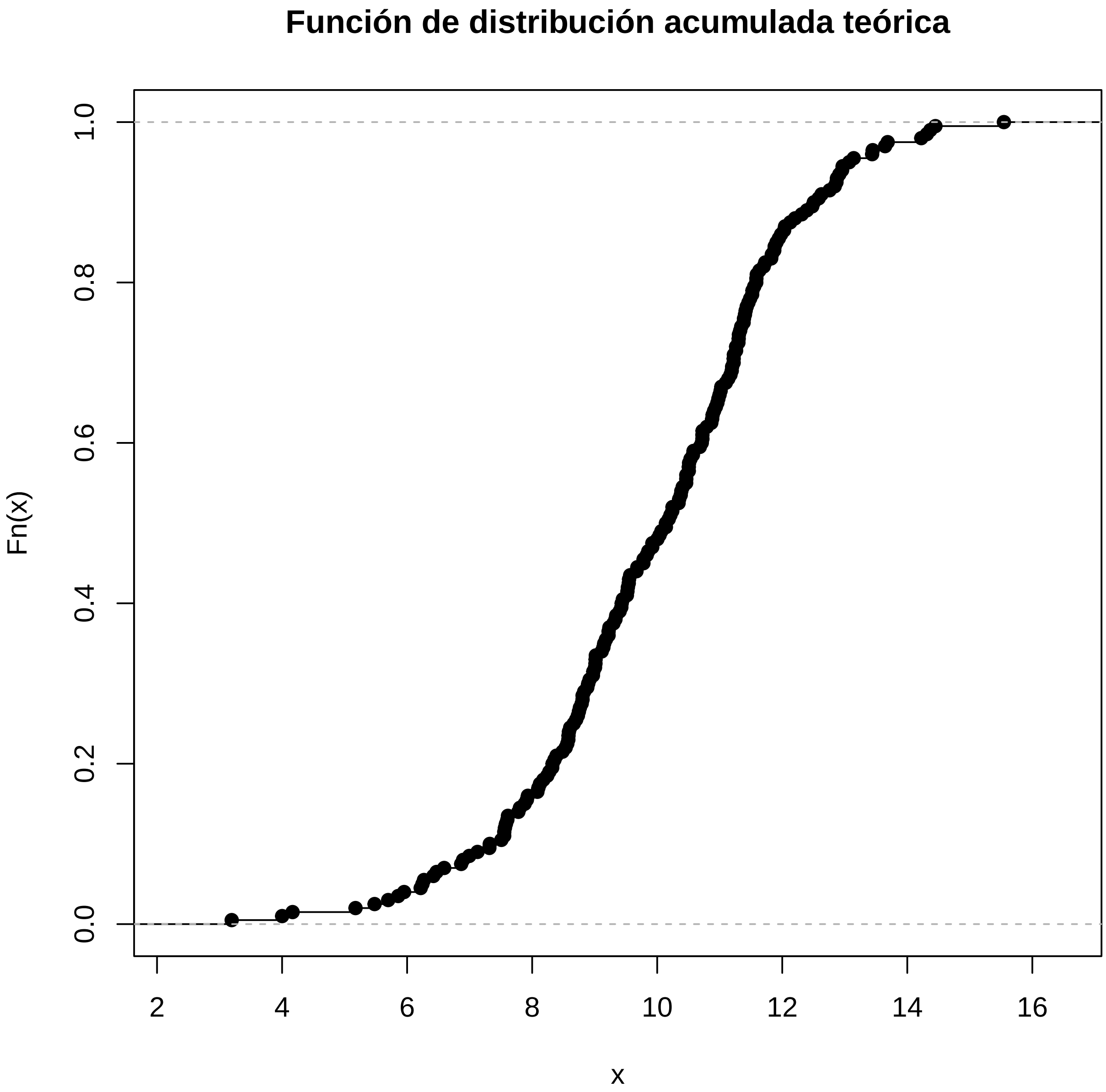
```
#genera los valores aleatorios de la distribución
x.norm <- rnorm(n=200,mean=10, sd=2)

# Podemos obtener un histograma usando la función hist()
hist(x.norm, breaks = "Sturges", freq = TRUE, probability = FALSE, include.lowest = TRUE,
     right = TRUE, density = NULL, angle = 45, col = "steelblue1", border = NULL,
     main = "Histograma de datos observados", axes = TRUE, plot = TRUE, labels = FALSE)
```




```
# Podemos estimar la densidad de frecuencia usando la función density() y plot()
#para dibujar su gráfica
plot(density(x.norm), main="Densidad estimada de los datos")
```




```
# R permite calcular la función de distribución acumulada teórica con ecdf()
plot(ecdf(x.norm),main="Función de distribución acumulada teórica")
```

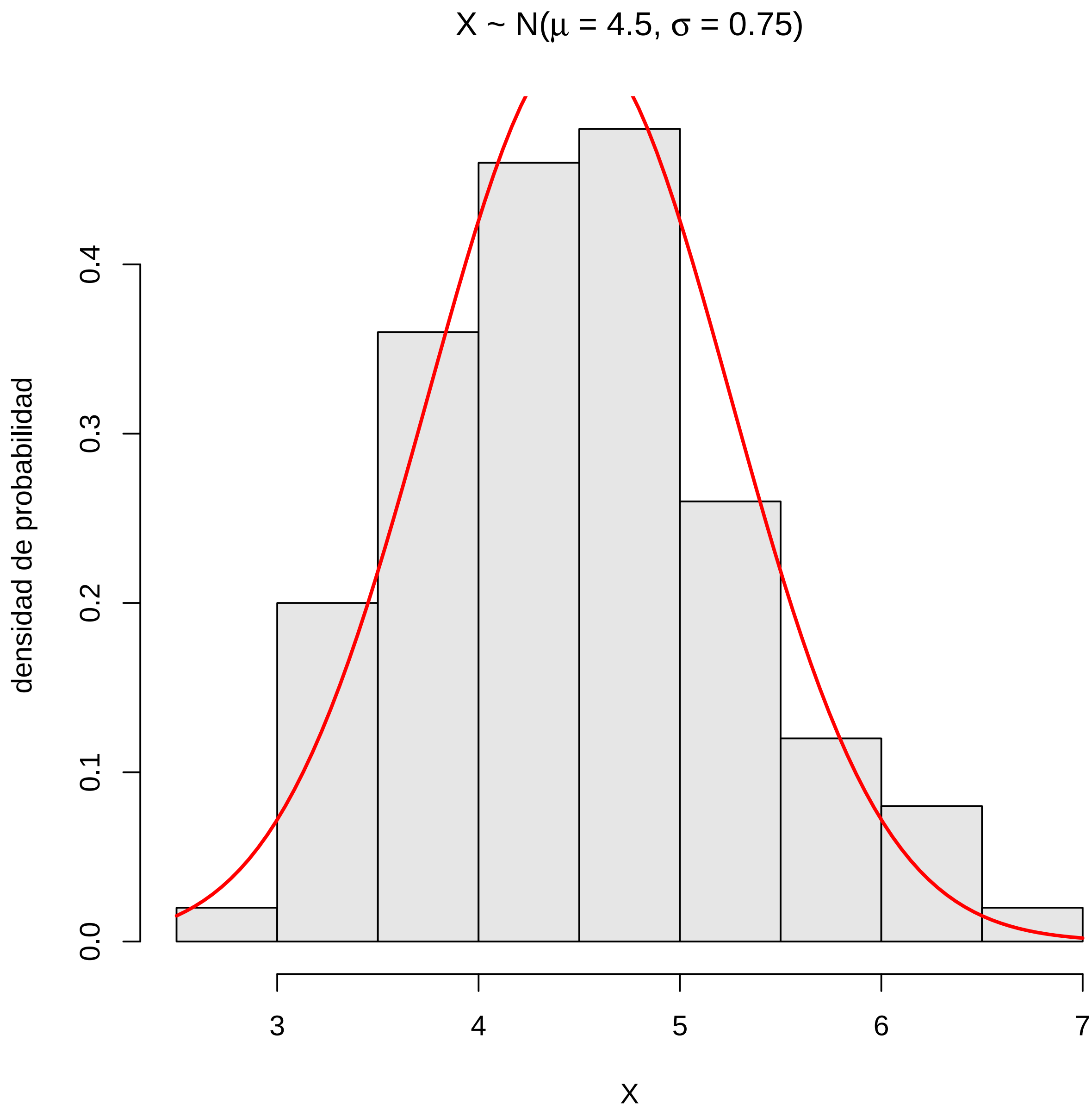


- **Ejemplo 3:**

Generar 100 números aleatorios de una distribución Normal con media 4.5 y desviación estándar 0.75

```
# Definir los parámetros apropiados
media <- 4.5
desviacion <- 0.75


# generar 100 números aleatorios de la distribución
x = rnorm(100, media, desviacion)
x

##   [1] 4.764416 4.039202 4.217986 6.271065 4.863098 5.667231 4.180365 4.693003
##   [9] 4.393026 4.978290 4.737783 3.841523 4.086329 2.940236 5.418882 3.689561
##  [17] 5.275740 3.850545 5.112313 4.906769 4.416071 4.575152 3.303397 5.541318
##  [25] 4.469147 4.352717 3.851215 4.787236 5.151771 3.759093 5.386894 3.245494
##  [33] 4.632891 4.068704 3.462969 6.167088 5.216762 4.277897 3.597141 3.994499
##  [41] 3.909655 4.221360 5.266447 3.708405 4.506891 3.399985 6.145905 4.888785
##  [49] 5.945433 4.803738 4.419292 3.971677 3.681422 6.138306 4.713995 5.271571
##  [57] 4.492637 5.158162 5.338901 4.443919 5.011008 4.041285 4.795439 4.238077
##  [65] 3.584627 4.003565 4.090517 3.811290 4.957908 3.085448 5.642024 4.562182
##  [73] 3.250636 4.500034 4.242403 5.738592 4.325977 4.581632 3.436834 3.072854
```

```
## [81] 4.846529 3.501944 4.558161 3.895118 3.492326 4.895469 4.726073 4.279570
## [89] 5.245092 4.239050 3.737393 4.554142 4.071045 3.815050 5.681962 5.176244
## [97] 3.947570 3.364465 6.806858 4.990999

# Histograma para la nuestra aleatoria de tamaño 100
hist(x,main=expression(paste("X ~ N(", mu, " = 4.5, ", sigma, " = 0.75)")),
     xlab="X", ylab="densidad de probabilidad", probability=TRUE, col=gray(0.9))


# Graficar la función de densidad teórica, usando la función curve()
curve(dnorm(x, media, desviacion), col="red", lwd=2, add=TRUE)
```

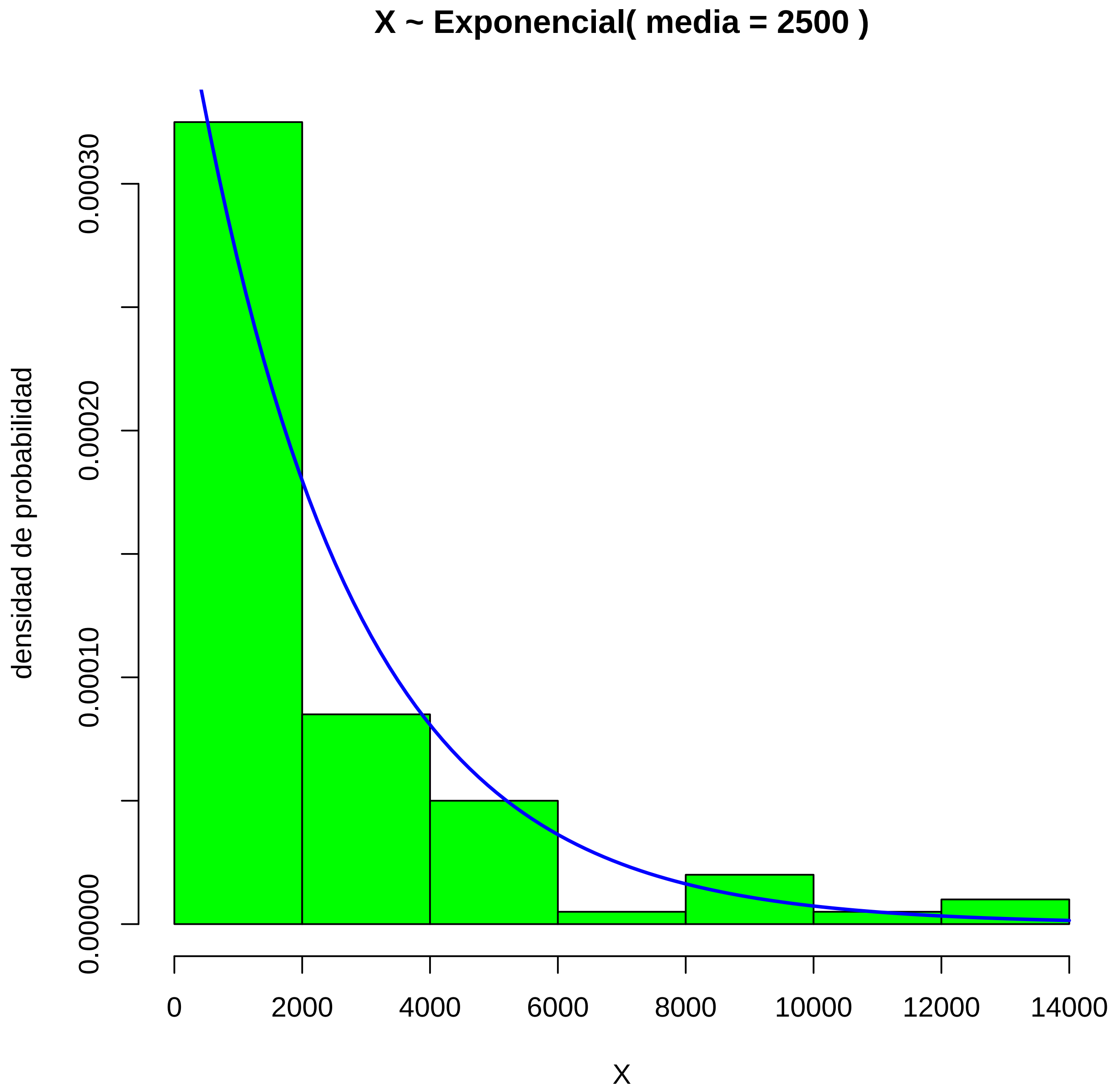


- **Ejemplo 4:**

Generar números aleatorios de una distribución exponencial. Por ejemplo, si la vida media de un bulbo de luz es 2500 horas, uno puede pensar que el tiempo de vida es aleatorio con una distribución exponencial que tiene media 2500. El único parámetro es la razón = 1/media.

```
# Definir el parámetro apropiado
media <- 2500
razon <- 1/media
n=100
```

```r
# generar 100 números aleatorios de la distribución
x = rexp(n, razon)
x

##   [1]  1616.68467  6665.02236  1498.95018  1232.77402  1883.91032  1593.16753
##   [7]  4359.66482  5105.09596  1312.81899  1575.51439  8446.80084  3917.05753
##  [13]   802.82885  4070.76289   618.00237   640.99011  2468.49094   406.76289
##  [19] 13554.67450   838.02011    60.33319  3987.00640   328.91058  2054.70171
##  [25]  2726.04527   515.76629  1793.29473   433.25578   927.98351   192.35576
##  [31]    80.08885   974.47768   385.62535  2669.25000  1425.90105   251.81763
##  [37]  3626.43777  1291.51916  1665.78243  1093.99692   998.19519  2913.76681
##  [43]   550.06666   247.59387   394.38550  9279.54530  2627.55659  1430.45445
##  [49]   327.45572   703.60212  1435.62258  5091.63024  3472.40794   928.95196
##  [55]   234.12319  1047.16130   745.81476  3391.03959  5815.57935  2273.30606
##  [61]  5192.42710 13614.93774   920.47692  8624.12121  1352.21456   797.03611
##  [67]   896.99695  1346.45355  2801.36848   916.64445   567.71183  1087.67996
##  [73]   795.80775  1673.07916   630.74929   267.96960  2378.17123  1913.92474
##  [79]  5258.69484  5432.17329   575.93819  1458.65204  5363.69929  4218.57139
##  [85]  1487.10982   763.69139  1232.42537 11320.22348  8439.34294   390.18016
##  [91]  1727.12337  3275.95507   995.82487   609.19832  1602.55002  3656.16538
##  [97]   759.48119  2568.19132  1464.28082  1634.17644

# Histograma para la nuestra aleatoria de tamaño 100
hist(x, main="X ~ Exponencial( media = 2500 )", xlab="X",
     ylab="densidad de probabilidad", probability=TRUE, col="green")


# Graficar la función de densidad, usando la función curve()
curve(dexp(x, razon), col="blue", lwd=2, add=TRUE)
```



4 FUNCIONES DE DISTRIBUCIÓN Y SU INVERSA (LOS CUANTILES).

En R, las funciones a las que se les antepone una "p" permiten contestar cuál es la probabilidad de que una variable aleatoria X sea menor o igual que x, esto es $F(x) = P[X \leq x]$. Las funciones a las que se les antepone una "q" son lo inverso de esto, ellas permiten conocer qué valor de una variable aleatoria X corresponde a una probabilidad p dada. Esto es el cuantil X_q o punto en el que los datos son partidos, $P[X \leq x_q] = p$

- **Ejemplo 1:**Para una Variable aleatoria X con distribución normal de media 1 y desviación estándar 1, ¿cuál es la probabilidad de que sea menor que 0.7?

```
x <- 0.7
p <- pnorm(x, mean=1, sd=1, lower.tail = TRUE)
p

## [1] 0.3820886
```

Observación: lower.tail=TRUE es el valor por defecto, para indicar las probabilidades son $P[X \leq x]$, en otro caso será $P[X > x]$.

- **Ejemplo 2:** Para una variable aleatoria con distribución normal estándar, encontrar $P[Z \leq 0.7]$ y $P[Z > 0.7]$.

```
z <- 0.7
p1 <- pnorm(z, mean=0, sd=1)
p1

## [1] 0.7580363

p2 <- pnorm(z, mean=0, sd=1, lower.tail=FALSE)
p2

## [1] 0.2419637
```

Observación: ya que $P[Z > 0.7] = 1 - P[Z \leq 0.7]$, obtenemos el mismo resultado con

```
p3 <- 1-pnorm(z, mean=0, sd=1)
p3

## [1] 0.2419637
```

- **Ejemplo 3:**¿Qué valor de una variable aleatoria con distribución normal estándar, tiene 75% del área a la izquierda?.

```
p <- 0.75
z <- qnorm(p, mean=0, sd=1, lower.tail = TRUE)
z

## [1] 0.6744898
```

Observación: note que el valor de z que resuelve $P[Z \leq z] = 0.75$ es el tercer cuartil (Q3), esto es $z = 0.6744898$

- **Ejemplo 4:**¿Cuál es la probabilidad a la derecha de 18.55 para una Variable aleatoria X con distribución Chi-cuadrado de 12 grados de libertad?

```
x <- 18.55
gl <- 12
p <- pchisq(x, gl, lower.tail = FALSE)
p

## [1] 0.09998251
```